

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-050295
 (43)Date of publication of application : 18.02.1997

(51)Int.Cl. G10L 5/04

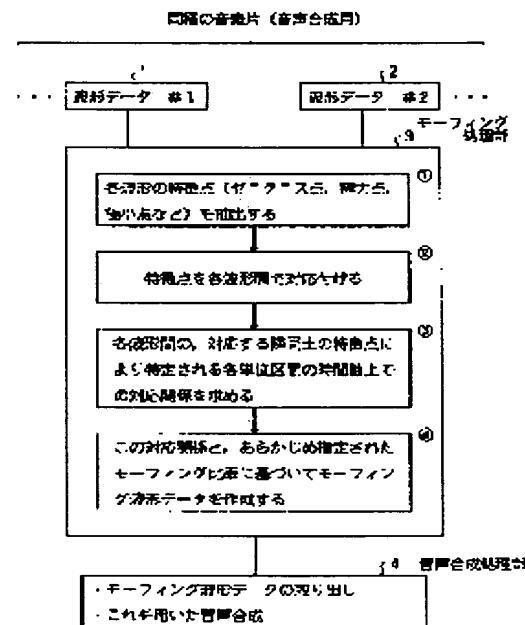
(21)Application number : 07-203168 (71)Applicant : FUJITSU LTD
 (22)Date of filing : 09.08.1995 (72)Inventor : OISHI KENICHI

(54) VOICE SYNTHETIC METHOD AND DEVICE THEREFOR

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a voice synthetic system high in convenience by facilitating the change of tone quality.

SOLUTION: Plural waveform data 1, 2... individually specifying respective waveforms related to the same kind of a phoneme piece are subjected to morphing, and a voice is synthesized by using the morphing waveform data subjected to the morphing. For instance, as the waveform data 1, the data taken out from a waveform dictionary in which the waveform data of respective phoneme pieces are previously held are used, and as the waveform data 2, the voice data sampled from the real (user's) voice data and extracted in phoneme piece are used. Also, a morphing ratio is adjustable.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

THIS PAGE BLANK (USPTO)

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-50295

(43)公開日 平成9年(1997)2月18日

(51)Int.Cl.⁶

G 10 L 5/04

識別記号

府内整理番号

F I

G 10 L 5/04

技術表示箇所

F

審査請求 未請求 請求項の数8 O.L (全9頁)

(21)出願番号

特願平7-203168

(22)出願日

平成7年(1995)8月9日

(71)出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72)発明者 大石 建一

神奈川県横浜市港北区新横浜2丁目4番19
号 株式会社富士通プログラム技研内

(74)代理人 弁理士 長谷川 文廣 (外2名)

(54)【発明の名称】 音声合成方法およびそのための装置

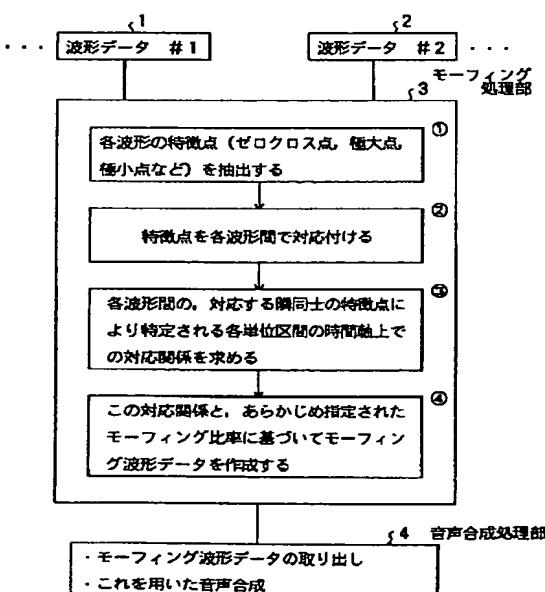
(57)【要約】

【課題】 音質変更を簡単に行なえるようにして利便性の高い音声合成システムを提供することを目的とする。

【解決手段】 同種の音素片についての各波形を個々に特定する複数の波形データ1、2・・・をモーフィング処理し、このモーフィング処理後のモーフィング波形データを用いて音声合成する。例えば波形データ1として、各音素片の波形データをあらかじめ保持している波形辞書から取り出したものを用い、波形データ2として、実際の(利用者の)音声データをサンプリングして音素片の単位で抽出したものを用いている。また、モーフィング比率は調整可能である。

本発明の基本構成図

同種の音素片(音声合成用)



【特許請求の範囲】

【請求項1】 同種の音素片についての各波形を個々に特定する複数の波形データをモーフィング処理し、このモーフィング処理後のモーフィング波形データを用いて音声合成するようにしたことを特徴とする音声合成方法。

【請求項2】 前記波形データの一つが、各音素片の波形データをあらかじめ保持している波形辞書から取り出したものであり、前記波形データの別の一つが、実際の音声データから音素片の単位で抽出したものであることを特徴とする請求項1記載の音声合成方法。

【請求項3】 前記モーフィング処理が、

- ・前記各波形のゼロクロス点、極大点、極小点などの特徴点を抽出し、
- ・当該特徴点を前記各波形間で対応付け、
- ・前記各波形間の、対応する隣同士の前記特徴点により特定される各単位区間の時間軸上での対応関係を求め、
- ・この対応関係と、あらかじめ指定されたモーフィング比率に基づいてモーフィング波形データを作成する、ものであることを特徴とする請求項1または2記載の音声合成方法。

【請求項4】 前記モーフィング比率が調整可能であることを特徴とする請求項3記載の音声合成方法。

【請求項5】 同種の音素片についての各波形を個々に特定する複数の波形データに対するモーフィング処理部と、

このモーフィング処理後のモーフィング波形データを用いて音声合成する音声合成処理部とを設けたことを特徴とする音声合成装置。

【請求項6】 前記波形データの一つが、各音素片の波形データをあらかじめ保持している波形辞書から取り出したものであり、

前記波形データの別の一つが、実際の音声データから音素片の単位で抽出したものであることを特徴とする請求項5記載の音声合成装置。

【請求項7】 前記モーフィング処理部での処理内容が、

- ・前記各波形のゼロクロス点、極大点、極小点などの特徴点を抽出し、
- ・当該特徴点を前記各波形間で対応付け、
- ・前記各波形間の、対応する隣同士の前記特徴点により特定される各単位区間の時間軸上での対応関係を求め、
- ・この対応関係と、あらかじめ指定されたモーフィング比率に基づいてモーフィング波形データを作成する、ものであることを特徴とする請求項5または6記載の音声合成装置。

【請求項8】 前記モーフィング比率を調整できることを特徴とする請求項7記載の音声合成装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声合成方法およびそのための装置に関し、特に同種の音素片についての各波形を個々に特定する複数の波形データをモーフィング処理し、このモーフィング処理後のモーフィング波形データを用いて音声合成する方式に関するものである。

【0002】近年のコンピュータシステムの一つの大きな流れであるマルチメディアにおいて、テキストベースの情報を音声情報に変換することができる音声合成処理は重要な機能を持つものであるが、この場合、音質変更の自由度が高くて利便性の優れた音声合成システムにすることが望ましく、本発明はこのような要請に応えるものである。

【0003】

【従来の技術】従来の音声合成処理の一つの手法である波形編集方式は、

- ・音声合成対象の元のテキスト（漢字かな混じり文）を言語解析してその中の各音素片およびそれぞれの時間長と抑揚パラメータなどを求め、
- ・この音素片の波形データを波形辞書から取り出し、
- ・この波形データを前記時間長や前記抑揚パラメータに基づいて調整する、

ようにしており、他の音声合成処理方式よりも音質面で優れていた。

【0004】なお、波形データは、ある人の音声から得られた音声データのうち任意の音素片ごとにその数周期分の波形を取り出したものであり、各音素片の波形データが波形辞書に格納されている。

【0005】そのため、波形辞書から得られる合成音声は、当然のことながらそこに格納されている波形データ生成元である特定人の声の特徴が強く残ったものとなり、合成音声の音質に変化を持たせたい場合には、この生成元が異なる種々の波形辞書（男女の区別、年令の違いなどに対応した各種の波形辞書）を用意しておき、これらの波形辞書を選択的に使用するといったことが行われている。

【0006】図6は、一般的な波形辞書のデータフォーマット例を示す説明図であり、波形辞書名21、音素片数22およびn個の音素片情報23を有しており、また音素片情報23のそれぞれは音素片名24と波形サンプリングデータ25とからなっている。

【0007】

【発明が解決しようとする課題】このように、従来の音声合成処理方式では、音質変更の自由度を確保するにはそれに見合うだけの波形辞書を用意しなければならないが、一般に、波形辞書の作成には多量の人手と時間がかかるため合成音声の音質を比較的自由に変えるようにすることはなかなか実現できないという問題点があつた。

【0008】そこで、本発明では、画像処理などの分野

で用いられているモーフィングの手法を音声合成の場合に適用することにより、音質変更を簡単に行なえるようにして利便性の高い音声合成システムを提供することを目的とする。

【0009】

【課題を解決するための手段】図1は、本発明の基本構成図である。図において、1は、モーフィング処理対象の波形データ（#1）であり、例えばあらかじめ用意されている波形辞書から取り出した音素片データである。2は、モーフィング処理対象の波形データ（#2）であり、例えば実際の（利用者の）音声データから音素片の単位で抽出したものであって、波形データ1とは同種の音素片が用いられる。3は、モーフィング処理部であり、モーフィング処理対象の各波形データ1、2からモーフィング波形データを作成し、これをメモリなど（図示省略）に格納している。4は、音声合成処理部であり、音声合成対象の各音素片に対応したモーフィング波形データを用いることにより合成音声を作成する。

【0010】なお、モーフィング処理の対象となる波形データ1、2……としてどの形式のものを用いるかは任意であり、例えば双方の波形データを、

- ・波形辞書から取り出す、
- ・実際の音声データから音素片の単位で抽出する、
- ようにもよい。

【0011】モーフィング処理部3における処理手順の概要は次のようにになっている（図2参照）。

- ①各波形のゼロクロス点、極大点、極小点などの特徴点を抽出する。
- ②当該特徴点を各波形間で対応付ける。
- ③各波形間の、対応する隣同士の前記特徴点により特定される各単位区間の時間軸上での対応関係を求める。
- ④この対応関係と、あらかじめ指定されたモーフィング比率に基づいてモーフィング波形データを作成する。なお、モーフィング比率自体は利用者が調整可能である。

【0012】ここで、特徴点の対応付けは、例えば各波形データ1、2の波形包絡線上のゼロクロス点、極大点、極小点などの特異点の中、時間軸上で逆転することなしに同じ態様で出現する同種のもの同士を組み合わせることにより行なわれる。

【0013】また、各波形包絡線上の前記単位区間に属する任意の点同士の時間的な対応関係は、この単位区間の時間軸上での当該各点の距離比を一致させることにより、確保される。

【0014】そして、モーフィング波形データは、例えば波形包絡線上の対応する点同士についての、

- ・単位区間の開始点からの両者の時間軸上の距離をモーフィング比率に基づく割合で加算した値の時間軸成分
- ・両者の振幅をモーフィング比率に基づく割合で加算した値の振幅成分

を持った新たな各点からなる波形（図2参照）を示すも

のである。

【0015】図2は、モーフィング処理前の波形（W、W'）と処理後のモーフィング波形（WM）とを示す説明図であり、このときのモーフィング比率は「50%・50%」に設定されている。

【0016】また、特徴点としては、

- ・そのレベルがマイナスからプラスに転じるゼロクロス点（A、A'）
- ・そのレベルがプラスからマイナスに転じるゼロクロス点（B、B'）
- ・△tの範囲内での極大点（C、C'）
- ・△tの範囲内での極大点（D、D'）

を用いている。

【0017】このようにして得られた各音素片のモーフィング波形データはメモリなどに格納されていわばモーフィング波形辞書を形成し、音声合成処理部は指定されたテキスト文（音声合成対象の漢字かな混じり文）の各音素片に対応するモーフィング波形データをこの波形辞書を順に取り出し、これらの波形データに基づく合成音声を生成する。

【0018】本発明の、音声合成方法としての基本的構成は、

「同種の音素片についての各波形を個々に特定する複数の波形データをモーフィング処理し、このモーフィング処理後のモーフィング波形データを用いて音声合成するようにしたこと」

である。

【0019】また、音声合成装置としての基本的構成は、

「同種の音素片についての各波形を個々に特定する複数の波形データに対するモーフィング処理部と、このモーフィング処理後のモーフィング波形データを用いて音声合成する音声合成処理部とを設けたこと」

である。

【0020】

【発明の実施の形態】図3～図5を参照して本発明の実施の形態を説明する。図3は、本発明の音声合成システムの全体構成を示す説明図であり、11はマイクロホン、12はプロセッサ、13は波形辞書、14はプログラムメモリ、15はマイクロホン11の出力である音声データ、16は波形辞書13から取り出した音素片データ、17は波形抽出処理の出力データである抽出波形データ、18はモーフィング処理の出力データであるモーフィング波形データをそれぞれ示している。

【0021】また、19は抽出波形データ17を保持する抽出波形保持部、20はモーフィング波形データ18を保持する波形メモリ、21はモーフィング比率、22は音声合成対象のテキスト文、23は音声合成処理の結果である合成音声をそれぞれ示している。

【0022】プログラムメモリ14には、

- ・サンプリング解析処理用プログラム
- ・波形抽出用プログラム
- ・モーフィング演算用プログラム
- ・音声合成用プログラム

などが保持されている。

【0023】また、モーフィング処理の対象データとして、
 ・マイクロホン11の出力である音声データ15をサンプリングおよび波形抽出することにより得られる抽出波形データ17
 ・あらかじめ波形辞書13に格納されている音素片データ16
 の両者を用いている。

【0024】この波形抽出において、子音の音素片では時間経過にともないその波形形状が変化して前後の音素片との境界が不明確となり抽出困難なケースがあるが、母音の音素片などでは抽出対象音声として単母音音節のものを用いれば、ほとんど波形形状の変化のない单一音素片の繰り返しとしての音声データが得られるため、その任意の波形周期を切り出すことにより当該音素片の抽出波形データ17が求まる。このようにして得られた抽出波形データ17は抽出波形保持部19に格納される。

【0025】続いて、波形辞書13と抽出波形保持部19からモーフィング対象の音素片データ16および抽出波形データ17を取り出して両者に対するモーフィング処理(図4および図5参照)を行ない、その実行結果であるモーフィング波形データ18が波形メモリ20に保持される。モーフィング比率はユーザが調整可能であり、この調整によって合成音声の音質を変えることができる。

【0026】続いて、指定されたテキスト文や制御信号に対する音声合成処理が波形メモリ20中のモーフィング波形データ18を用いて行なわれ、その実行結果が合成音声として用いられることになる。

【0027】なお、前記の子音のように、テキスト文の音素片の中で対応するモーフィング波形データ18が波形メモリ20から得られないものについては、波形辞書13の音素片データ16がそのまま合成音声要素として用いられる。

【0028】上述の実施例では、モーフィング処理の元となる一方の波形データとして、マイクロホン11の出力である音声データ15をサンプリングおよび波形抽出することにより得られる抽出波形データ17を用いているが、これに代えて波形辞書13とは性質の異なる別の波形辞書から取り出した音素片データを用いるようにしてもよい。

【0029】図4および図5は、二つの波形データに対するモーフィング処理の手順を示す説明図であり、その内容は次のようにになっている(図2参照)。なお、以下の説明では二つの波形データをそれぞれW, W'、また時間tにおけるそれぞれの振幅をL(t), L'(t)

と記載する。

【0030】(11)WおよびW'について、次の特徴点設定処理を行う(図2参照)。

- ・そのレベルがマイナスからプラスに転じるゼロクロス点を特徴点A, A'に設定
- ・そのレベルがプラスからマイナスに転じるゼロクロス点を特徴点B, B'に設定
- ・その時間の前後△tの範囲でレベルが極大となる点を特徴点C, C'に設定
- ・その時間の前後△tの範囲でレベルが極小となる点を特徴点D, D'に設定

【0031】続いて、両波形データの対応区間の設定処理を実行する。

(12)WおよびW'それぞれの特徴点A, A'を探す。

(13)特徴点A, A'の個数が同じであるかどうかを判断し、その結果が「YES」の場合は次のステップに進み、「NO」の場合はモーフィング処理を終了する。

(14)ステップ(12)で求めた各特徴点A, A'についての時間軸上での写像f_Aを定める。

(15)開始点S, S'、特徴点A, A'及び終了点E, E'の各点で分割した、両波形の区間同士を対応させる。すなわち、図2のWについては、「S～A₁」、「A₁～A₂」および「A₂～E」の3区間に分割し、例えばこの中の「A₁～A₂」をW'の「A₁'～A₂'」に対応させる。

(16)ステップ(15)で求めた区間ごとに特徴点B, B'を探す。

(17)対応する当該区間ごとの特徴点B, B'の個数が同じであるかどうかを判断し、「YES」の場合は次のステップに進み、「NO」の場合はモーフィング処理を終了する。

(18)ステップ(16)で求めた各特徴点B, B'についての時間軸上での写像f_Bを定める。

(19)開始点S, S'、特徴点A, A'、特徴点B, B'及び終了点E, E'の各点で分割した、両波形の区間同士を対応させる。すなわち、図2のWについては、「S～A₁」、「A₁～B₁」、「B₁～A₂」および「A₂～E」の4区間に分割し、例えばこの中の「A₁～B₁」をW'の「A₁'～B₁'」に対応させる。

(20)ステップ(19)で求めた区間ごとに特徴点C, C'を求める。

(21)対応する当該区間ごとの特徴点C, C'の個数が同じであるかどうかを判断して、「YES」の場合はステップ(23)に進み、「NO」の場合は次のステップに進む。

(22)余剰特徴点を削除する。

(23)ステップ(20)で求めた各特徴点C, C'についての時間軸上での写像f_Cを定める。

(24)開始点S, S'、特徴点A, A'、特徴点B, B'、特徴点C, C'及び終了点E, E'の各点で分割した、両波形の区間同士を対応させる。すなわち、図2

のWについては、「S～A₁」、「A₁～C₁」、「C₁～C₂」、「C₂～C₃」、「C₃～B₁」、「B₁～C₄」、「C₄～C₅」、「C₅～A₂」および「A₂～E」の9区間に分割して、例えばこの中の「C₁～C₂」をW'の「C₁'～C₂'」に対応させる。

(25)ステップ(24)で求めた区間ごとに特徴点D, D'を求める。

(26)対応する当該区間ごとの特徴点D, D'の個数が同じであるかどうかを判断して、「YES」の場合はステップ(28)に進み、「NO」の場合は次のステップに進む。

(27)余剰特徴点を削除する。

(28)ステップ(25)で求めた各特徴点D, D'についての時間軸上での写像f_Dを定める。

(29)開始点S, S'、特徴点A, A'、特徴点B, B'、特徴点C, C'、特徴点D, D'及び終了点E, E'の各点で分割した、両波形の区間同士を対応させる。すなわち、図2のWについては、「S～A₁」、「A₁～C₁」、「C₁～D₁」、「D₁～C₂」、「C₂～D₂」、「D₂～C₃」、「C₃～B₁」、「B₁～D₃」、「D₃～C₄」、「C₄～D₄」、「D₄～C₅」、「C₅～D₅」、「D₅～A₂」および「A₂～E」の14区間に分割し、例えばこの中の「C₂～D₂」をW'の「C₂'～D₂'」に対応させる。

【0032】続いて、両波形上の任意の点のについての時間軸上での写像fを定める処理を実行する。

(30)W上の任意の点Pの、それが属している区間Kの時間幅Tに対する内分比rを求める。

(31)W'上の点であって、区間Kに対応の区間K'の時間幅T'を当該内分比rで分割したときの点P'を求める。

(32)所定数の分割処理が終了したかどうかを判断し、「YES」の場合は次のステップに進み、「NO」の場合はステップ(30)に戻る。

(33)W上の点PとW'上の点P'との対応関係に基づいて、両波形データの時間軸上での写像fを定める。

【0033】続いて、WおよびW'の波形データに対するモーフィング処理を実行する。

(34)次式の計算を実行してモーフィング波形データWM(t)を求める。このときのモーフィング比率は[m : n]であり、モーフィング波形データの時間tにおける振幅は、tの時点でのWのそれとf(t)の時点でのW'のそれを当該モーフィング比率で内分した値になっている。

【0034】

【数1】

$$WM(t) = \frac{[m \times L(t) + n \times L'(f(t))]}{m+n}$$

(m, nはモーフィング比率)

【0035】ステップ(14)、(18)、(23)およびステップ(28)のそれぞれで定まる写像はモーフィング波形データWM(t)の算出には利用されず、当該ステップを実行するかどうかは任意である。

【0036】ステップ(15)、(19)、(24)およびステップ(29)における両波形間での区間同士の対応付けの処理に際しては、当該区間を特定する各特徴点の時間情報が用いられる。

【0037】

【発明の効果】このように、本発明では、音声合成時の波形データとして、モーフィング処理により生成したモーフィング波形データを用いているので、従来のように種別の異なる多数の波形辞書を用意しなくても、モーフィング処理対象のデータやモーフィング比率などを変えることにより多くのバリエーションの波形データが得られる。

【0038】そのため、音声合成に際しての音質変更の自由度が高く利便性に優れた音声合成方法およびそのための装置を提供することができ、また音声合成システムとしての性能向上を図ることができる。

【図面の簡単な説明】

【図1】本発明の、基本構成図である。

【図2】本発明の、モーフィング処理前の波形(W, W')と処理後のモーフィング波形(MW)を示す説明図である。

【図3】本発明の、音声合成システムの全体構成を示す説明図である。

【図4】本発明の、波形データW, W'に対するモーフィング処理の手順を示す説明図(その1)である。

【図5】本発明の、波形データW, W'に対するモーフィング処理の手順を示す説明図(その2)である。

【図6】一般的な、波形辞書のデータフォーマット例を示す説明図である。

【符号の説明】

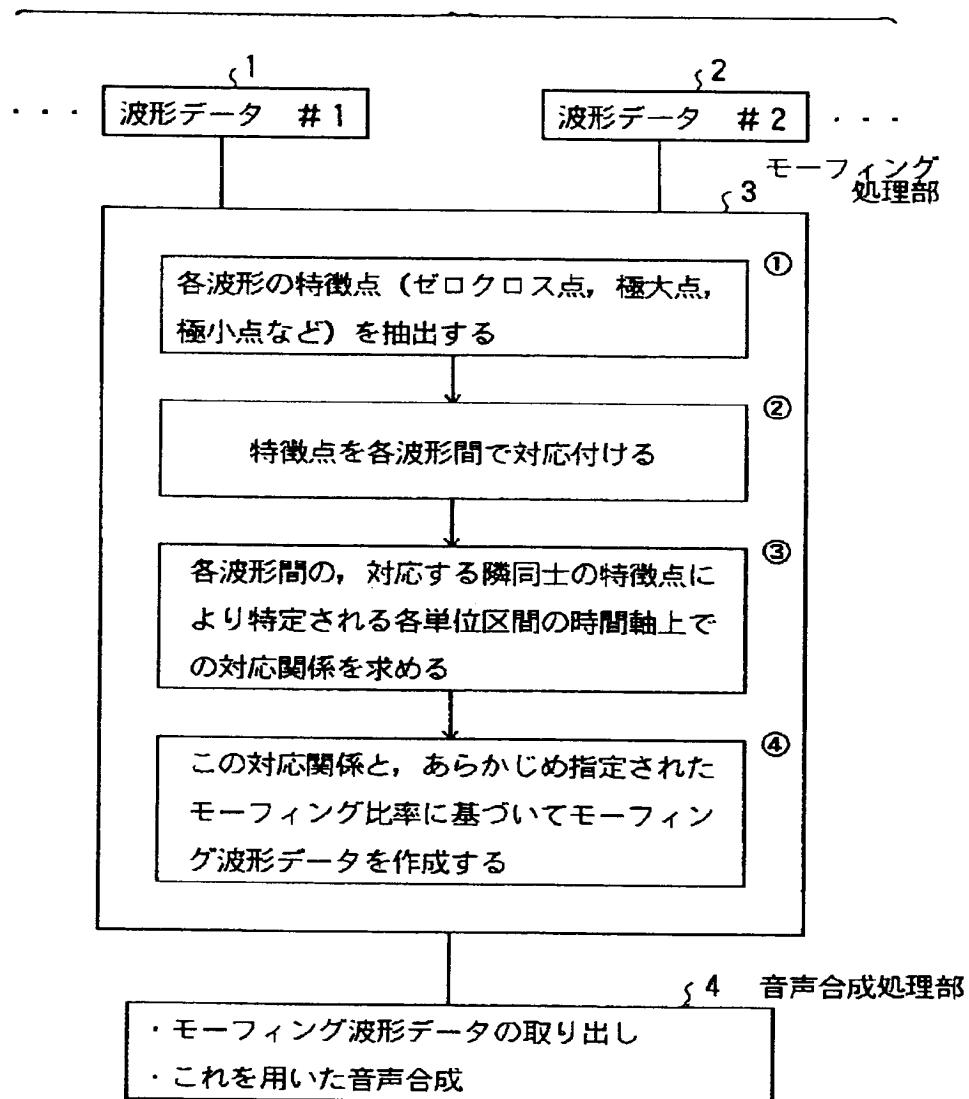
図1において、

- 1 … 波形データ(♯1)
- 2 … 波形データ(♯2)
- 3 … モーフィング処理部
- 4 … 音声合成処理部

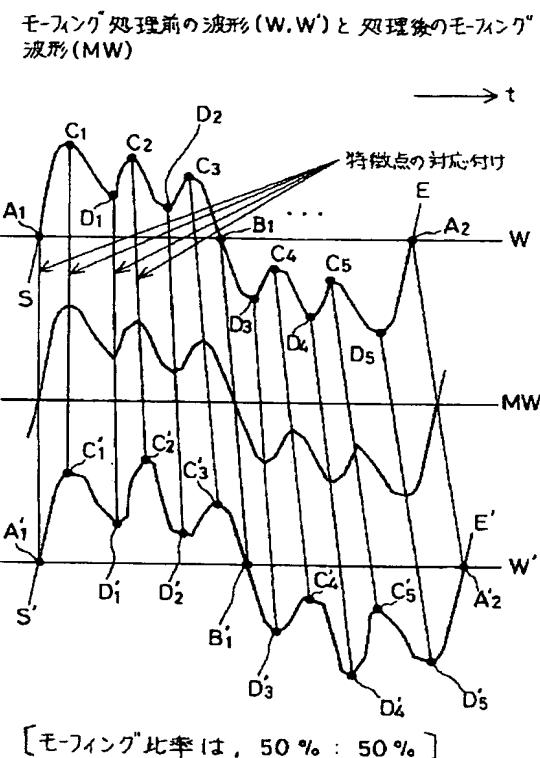
【図1】

本発明の基本構成図

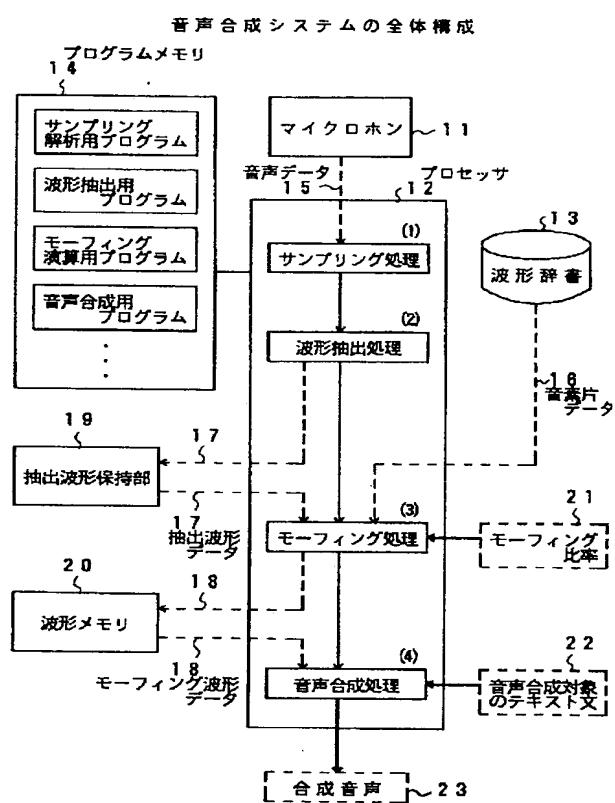
同種の音素片（音声合成用）



【図2】



【図3】

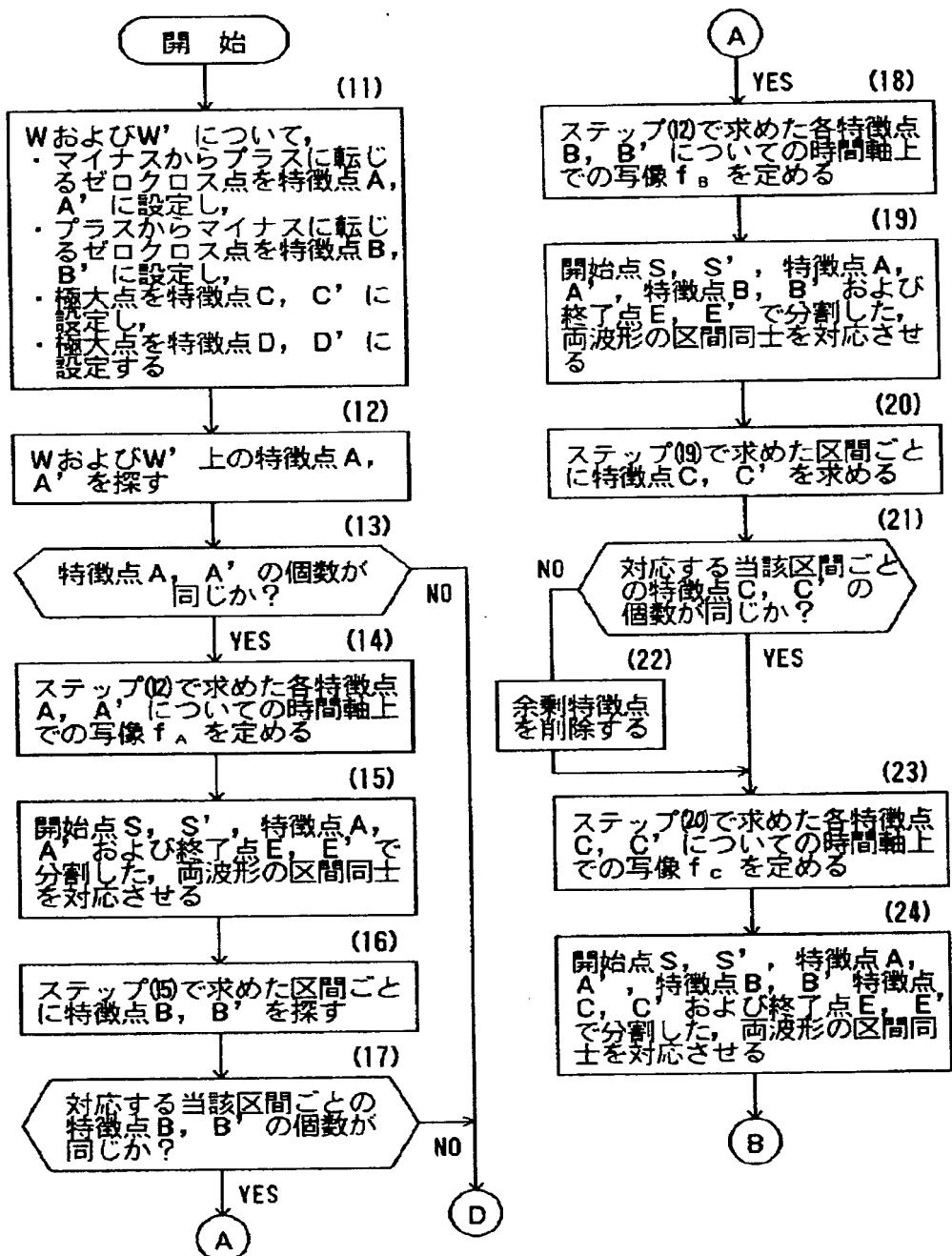


【図6】

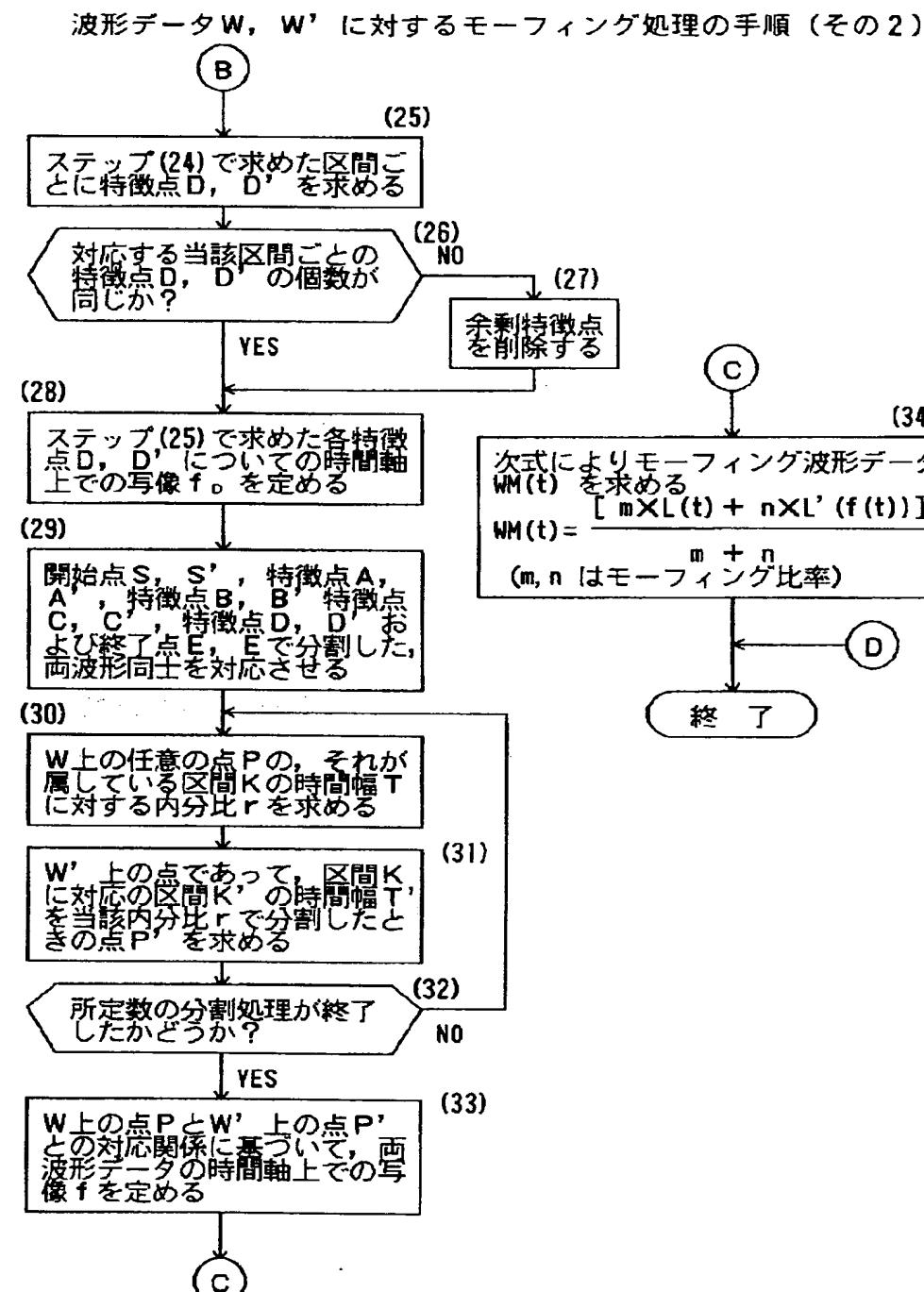
一般的な、波形辞書のデータフォーマット例

波形辞書名	～21
音素片数	～22
音素片情報 [0]	～23
音素片名	～24
波形サンプリングデータ (数周期分 [数 msec 分])	～25
音素片情報 [1]	～23
音素片名	～24
波形サンプリングデータ	～25
⋮	⋮
音素片情報 [音素片数-1]	～23
音素片名	～24
波形サンプリングデータ	～25

【図4】

波形データ W, W' に対するモーフィング処理の手順 (その1)

【図5】



THIS PAGE BLANK (USPTO)